

³Center for General Linguistics, Schuetzenstrasse 18, Berlin, 10117, Germany

INTRODUCTION

Inner speech is an intriguing phenomenon that occurs during activities such as mind wandering, problem solving, planning, reading, and writing. It reflects the perception of an inner voice that is physiologically attributed to the self. There is considerable debate on whether inner speech is abstract [Jones, 2009; Oppenheim and Dell, 2008, 2010] or whether it is a more concrete phenomenon with representation of phonetic detail that involves a motor simulation [Corley et al., 2011]. Theoretical proposals reconciled these views by suggesting different levels of abstraction from abstract inner speech during mind wandering to concrete inner speech during reading and writing [for a review, see Perrone-Bertolotti et al., 2014]. The concreteness of inner speech during reading is thought to increase when inner speech is mouthed during silent articulation [subvocal speech, Conrad and Schönle, 1979]. After an initial phase of overt reading during reading acquisition, children often mouth while they read silently. This suggests that inner speech could aid in acquiring reading skills. In adults, inner speech supports reading by implicitly assigning prosodic structure to read sentences [Kentner and Vasissth, 2016].

Inner speech has been investigated in various functional imaging paradigms, including covert naming, working memory, and reading. However, covertness was frequently used to preclude motion artifacts rather than constituting the research interest [De Nil et al., 2000; Gruber and von Cramon, 2003; Koelsch et al., 2009; Mazoyer et al., 2016; for a review see Alderson-Day and Fernyhough, 2015]. So far, imaging studies on voice-related aspects of subvocal processing focused on the role of inner speech in emulating prosody. The left hemisphere of the brain is sensitive to prosodic modulations affecting syntax during covert sentence reading [Loevenbruck et al., 2005]; and voice-sensitive regions in the right more than left auditory association cortex contribute to the generation of inner speech [Yao et al., 2011]. Reading-related activity in these regions occurs relatively late after stimulus onset and is modulated by attention and thus does not simply constitute a consequence of multisensory bottom-up processing [Perrone-Bertolotti et al., 2012]. Rather, it was interpreted as simulation of prosody during inner speech [Perrone-Bertolotti et al., 2014]. Yet, inner speech may contribute more to reading than just speech melody. It could potentially assist in grapho-phonological transformations underlying reading, a function that has been associated with left-lateralized processing in reading models [Binder et al., 2009; Jobard et al., 2003; Price, 2012; Taylor et al., 2013]. This would require a fine enough resolution of inner speech that provides sufficient phonological detail. The degree of concreteness of inner speech during reading has not yet been investigated.

Using fMRI, we tested whether the brain, particularly the left hemisphere, codes phonetic detail during inner speech while reading covertly. We compared two inner (covert reading and silently articulating read sentences)

with two overt speech conditions (listening to and overtly reading the same sentences). We have two main assumptions: first, fMRI identifies phonetic detail during overt speech processing on the neural level; second, a comparable phonetic detail is expected to be observed during inner speech conditions even if no auditory input reaches the brain. Although phonetic detail is coded in acoustic changes that are by themselves much faster than fMRI time resolution, there is accumulating evidence that manipulations of auditory temporal modulation rates of speech or speech-like stimuli induce changes in temporal lobe activity that can be measured with fMRI [Arnold et al., 2013; Arsenault and Buchsbaum, 2015; Baumann et al., 2015; Boemio et al., 2005; Jancke et al., 2002; Pichon and Kell, 2013]. These studies do not imply that BOLD captures these neural dynamics directly but they suggest rather that BOLD identifies brain regions that are sensitive to such a modulation. These may include early auditory cortices as well as downstream regions.

Our findings are closely tied to two influential neuro-computational models of speech production, the DIVA [Tourville and Guenther, 2011] and the hierarchical state feedback model [Hickok, 2012]. The former represents the most detailed model of overt speech production. It proposes a left frontal feedforward control system together with feedback loops for integration of external auditory and somatosensory feedback involving bilateral auditory and somatosensory cortices and the right ventral premotor cortex. As such, it provides testable predictions for the processing of external auditory and somatosensory feedback during overt speech. Yet, the DIVA model does not provide clear predictions for covert speech processes. The hierarchical state feedback model fills this gap by specifying internal frontotemporal loops that could aid in speech planning and inner speech. We hypothesized that the inner speech conditions exhibit phonetic detail comparable to those observed under the overt speech conditions decipherable in frontal and/or temporal cortices given the emulation of an inner voice. Should inner speech indeed provide such detail, then voicing effects down to the level of consonant features should be detectable.

Consonants differ primarily along three major axes: place of articulation, manner of articulation, and voicing [Pickett, 1999]. They reflect the way the consonant is produced by the speech production system. The place of articulation points to the articulatory zone in which the constriction is realized. For German, the language of investigation in this study, it ranges from bilabial to glottal [see Kohler, 1990 for an illustration of the German consonant inventory]. Processing of place of articulation during speech perception has been connected to the left-lateralized dorsal speech stream [Arsenault and Buchsbaum, 2015; Murakami et al., 2015] and can be decoded from the articulatory motor cortex during overt speech production based on its somatotopic organization [Bouchard et al., 2013]. The manner of articulation describes

the way articulators are modified to produce the airflow used to make a speech sound [Hall, 2000]. In German, five different manners of articulation can be differentiated: stops, nasals, fricatives, approximants and lateral approximants [Kohler, 1990]. Stops belong to the major class of obstruents, while nasals, approximants, and lateral approximants belong to sonorants. The key difference between sonorants and obstruents is connected to the third consonantal feature, voicing, which is the phonetic detail investigated in this study. Voicing characterizes the vibration of the vocal folds, in contrast to voicelessness, where vocal folds' vibration is absent. While sonorant consonants are always phonologically voiced, obstruents can be either phonologically voiced or voiceless. The German obstruent inventory reveals a relatively equal distribution of voiced and voiceless stops [Kohler, 1990; see also Marschall 1995]. Voicing is of particular interest in this study, because it does not only separate minimal pairs in the class of obstruents, but additionally contains prosodic information that pertains to the metrical (stressed and unstressed syllables) and information structures (accented or not) as well as the speaker's communicative intention (e.g., question or statement) in larger temporal windows than that of a phoneme [Nooteboom, 1997].

Phonation in phonologically voiced obstruents is a result of stretched vibrating vocal folds and a transglottal pressure difference between the pressure below and above the folds [Titze et al., 1995]. Phonologically voiceless obstruents occur without phonation and with an open glottis. Thus, both voiced and voiceless phonemes involve laryngeal adjustments. This may explain the failure to decode voicing from articulatory motor cortex during articulation [Bouchard et al., 2013].

Yet, voicing has major acoustic consequences. It drastically affects acoustic amplitude and thus the power envelope of the perceived speech signal (for representative examples, see Fig. 1). There is a close relationship between temporal acoustic speech features and temporal modulation of brain activity during speech processing. The auditory cortex tracks the acoustic power envelope during speech perception [Gross et al., 2013; Luo and Poeppel, 2007]. Consequently, voicing has been identified as a major feature of auditory cortex organization [Mesgarani et al., 2014; Nourski et al., 2015]. Should inner speech be concrete enough to represent the presence or absence of voicing, activation differences could potentially be detected in auditory association cortices that represent the emulated auditory targets of inner speech [Hickok, 2012]. Amplitude is not the only characteristic of a given speech signal affected by voicing, as frequency composition is influenced, too. Voiceless consonants lack the fundamental frequency (the frequency of vocal fold vibrations), which results in drastically reduced power in a rather low frequency range (below 400 Hz) compared with voiced consonants. Voiceless fricatives provide high energy in very high acoustic frequencies up to 8 kHz [Ladefoged and

Maddieson, 1996]. The lack of fundamental frequency and the higher frequency components of fricatives result in overall higher relative acoustic frequencies of voiceless compared with voiced consonants.

A theoretical framework for speech perception, which focuses on the frequency domain, provides a testable hypothesis for the lateralization of consonant processing also applicable to the inner speech conditions. The Double Filtering by Frequency Theory posits that the left hemisphere preferentially processes higher frequency components of the speech spectrum [Ivry and Robertson, 1999]. Ivry and Robertson explicitly formulate the prediction that the processing of voiceless consonants should lateralize more strongly to the left than the processing of voiced consonants given their overall higher spectral frequency content. Indeed, identification of initial voiceless consonants embedded in auditory consonant-vowel (CV) syllables elicits stronger BOLD activation of the left auditory cortex compared with identification of CV syllables with initial voiced consonants [Jancke et al., 2002]. In contrast, voice-selective regions that should increase their activity with the amount of voicing based on amplitude changes have been identified bilaterally in the superior temporal sulcus (STS) [Belin et al., 2000].

In this study, we tested whether inner speech represents detail as comprehensively as consonant voicing, even in the absence of auditory input. In the positive case, sentences with more voiced consonants and henceforth voicing should activate voice-sensitive regions in the bilateral STS more strongly than sentences with more voiceless consonants even in the inner speech conditions. Sentences with more voiceless consonants should, in turn, evoke stronger left-lateralized activity in phonological regions that process relative higher spectral frequencies. Of note, both the DIVA and the hierarchical state feedback control models so far are unidimensional in the sense that they do not propose different subregions for the processing of the various acoustic dimensions.

The critical phonological manipulation in this experiment consisted in the variation of the amount of voiceless and voiced consonants in syntactically identical, semantically nonsense German sentences. Sentences were presented in randomized order in every aforementioned condition. By testing for a positive and negative parametric modulation of task-related activity associated with consonant voicing and entering the resulting beta values in a conjunction analysis across tasks, we targeted voicing correlates that occurred independently of condition. These activations should point to brain regions that are sensitive to consonant features during sentence processing even in the inner speech conditions. The general assumption here is that those brain regions are sensitive to both sensory input and predictions of internal forward models [Hickok, 2012].

Condition contrasts independent of voicing additionally allowed for investigation of lateralization effects associated

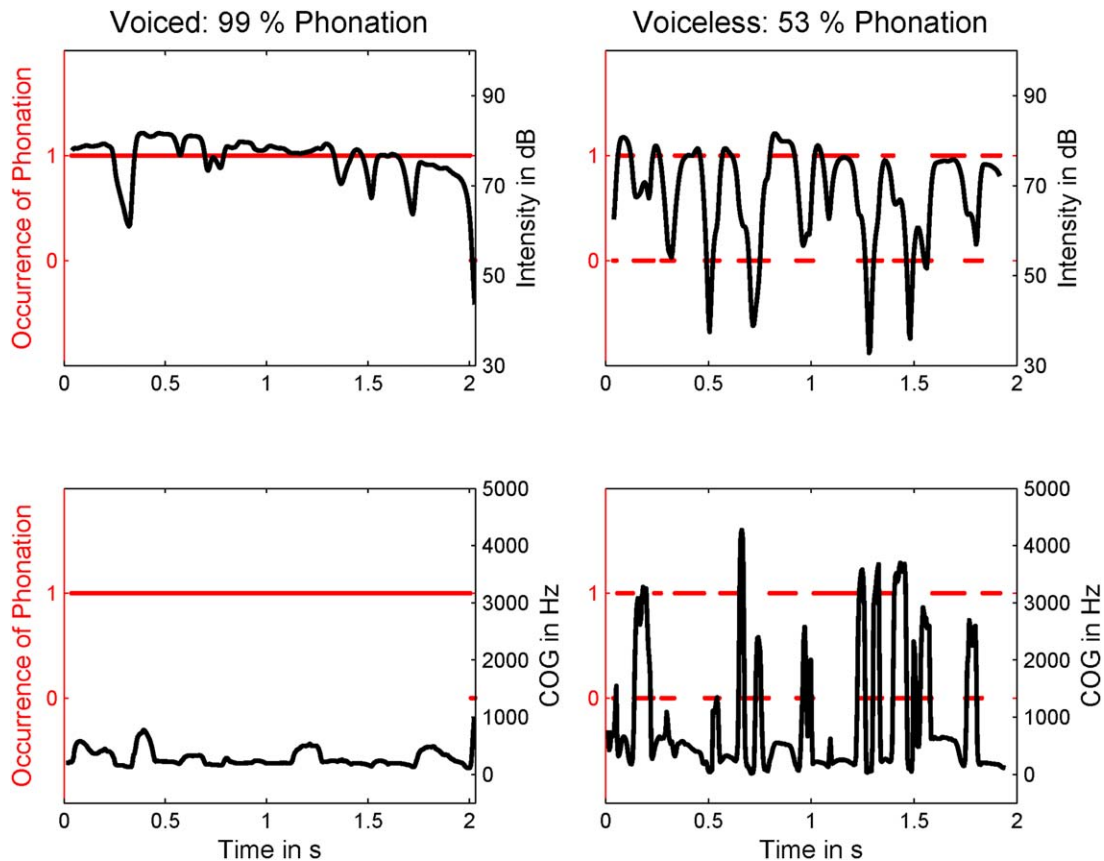


Figure 1.

Representative example of the acoustic measures obtained from a sentence rich in voiced (left panels) and a sentence rich in voiceless consonants (right panels) of the male speaker. The presence or absence of phonation (voicing) as measured as fundamental frequency in PRAAT is displayed as 1 or 0 (red horizontal lines, left red y-axis). The overall amount of phonation is given as a

percentage on top of the two upper panels. Changes in Intensity (upper panels) and spectral Center of Gravity (COG, lower panels) during the sentence are displayed in black (right y-axes). The x-axes correspond to Time in seconds. [Color figure can be viewed at wileyonlinelibrary.com.]

with somatosensory and auditory feedback processing during articulation. We expected temporal activations associated with auditory feedback when contrasting overt reading with covert reading while mouthing the sentences. Comparing overt reading with listening to the same sentences should reveal brain regions that are sensitive to auditory feedback of the self-produced speech (auditory error maps in the DIVA and auditory targets in the state feedback control model). Self-generated speech compared with listening to pre-recorded speech causes a relative suppression of bilateral anterior superior temporal activations and enhances bilateral posterior superior temporal cortex activity [Agnew et al., 2013]. Because online corrections of auditory or somatosensory feedback perturbations activated right-lateralized premotor cortices in addition to the bilateral auditory or somatosensory cortex activations, the DIVA model proposes a right-lateralized error monitoring system for speech production [Golfopoulos et al., 2011;

Tourville et al., 2008]. Such right-lateralization of sensory feedback processing is challenged by the observation of bilateral sensorimotor integration during speech production both on the BOLD and the neural level [Agnew et al., 2013; Cogan et al., 2014; Dhanjal et al., 2008; Simmonds et al., 2014]. Sensorimotor integration during speech processing has also been proposed to occur in the left parieto-temporal junction [Buchsbaum et al., 2011; Hickok et al., 2003, 2009; Hickok and Poeppel, 2007]. Overt compared with covert reading showed left-lateralized activity in this region [Kell et al., 2011]. In sum, even lateralization of auditory and somatosensory feedback processing during speaking is still a matter of debate. Thus, we additionally investigated lateralization of condition contrasts of listening to sentences (auditory input), reading sentences covertly (visual input, inner speech), reading sentences while mouthing (articulating them without phonating; visual input, more concrete inner speech, sensorimotor processing), and reading sentences overtly (visual input, sensorimotor

TABLE I. Stimulus material, sorted by the number of voiced consonants

Weise Marder lügen nie in den Saal
 Arge Mäuler lügen gern in der Bar
 Müde Leiber baden lange im Moor
 Böse Buben jagen übel den Gaul
 Maue Bande lümmeln lahm um den Ball
 Wunde Daumen wollen wieder ihr Wohl
 Gelbe Waben bilden neu einen Bau
 Lila Blumen blühen wohl an der Alm
 Lahme Maden mümmeln öde den Brei
 Blaue Meisen loben weise den Wurm
 Milde Laugen binden vage den Leim
 Liebe Löwen saugen nur an dem Bein
 Lose Garne weben silbern den Saum
 Wollne Bündel wabern bleiern im Darm
 Geile Wagen biegen doll in die Bahn
 Lange Dünen liegen gülden am Meer
 Dünne Geigen weinen leise in Moll
 Alle Bühnen geben nun einen Ball
 Saure Sahnen laben länger den Wal
 Blöde Weiber glauben blind jeden Wahn
 Linde Damen glühen liebend beim Wein
 Grüne Algen düngen milde den Wald
 Warme Wellen wiegen weich um dem Wall
 Laue Winde wedeln gnädig dem Sohn
 Wahre Wonnen lodern siedend wie Öl
 Beide Leiber lieben glühend am Main
 Weise Ahnen dulden gnädig ihr Sein
 Dumme Wale mögen niemals den Aal
 Noble Gaben landen immer im All
 Miese Bengel mögen alle nur Gold
 Leise Mägde buhlen albern im Mai
 Böse Mäuse nagen bange am Laub
 Güldne Dolden wogen blühend am Weg
 Blonde Haare wallen golden im Wald
 Runde Münder singen gläubig ein Lied
 Neue Blusen hängen wiegend am Seil
 Blinde Eulen leiern duldend ihr Leid
 Hohle Wege lähmen dauernd wie Blei
 Beige Wannen laden lau in ein Bad
 Wilde Geier lauern äugend im Baum
 Tätige Affen kicken zu dem Pott
 Krasse Kassetten husten in den Strauch
 Feuchte Zehen tasten nach den Kassen
 Kranke Schichten achten nicht die Hütte
 Falsche Kissen heizen oft die Schule
 Kesse Autos pfeifen durch die Straßen
 Heiße Kisten schenken fast einen Kuss
 Starke Zeichen speichern hier die Taschen
 Fette Physiker passen durch das Fach
 Keusche Effekte schauen noch den Sport
 Schattige Stufen stoppen spät den Hirsch
 Leckere Kaffees kosten nicht viel Zeit
 Taffe Schäfte peitschen sonst die Pfeifen
 Echte Köche spuken den letzten Takt
 Putzige Kappen hüpfen oft im Schuh
 Fette Töpfe backen oft den Zucker
 Frische Kirschen stoßen tüchtig die Frucht
 Faktische Stützen hacken seit dem Fest

TABLE I. (continued).

Kurze Schiffe stoppen plötzlich den Staat
 Falsche Zicken küssen heute am Stück
 Höfliche Kühe fechten fast den Tag
 Optische Schlösser züchten auf dem Schluss
 Tiefe Kutschen schießen rechts der Katze
 Pfiffige Füße kochen jetzt am Spieß
 Hektische Fische hocken auf dem Eis
 Aktive Teiche putzen die Zeche
 Spitze Stifte tauschen hier den Schützen
 Schiefe Kirchen tauchen zurück zum Schacht
 Heftige Stiche essen oft Kuchen
 Feiste Texte stechen nicht den Kaktus
 Skeptische Töchter packen jetzt den Schatz
 Fitte Fässer zapfen stets das Schnäppchen
 Hohe Texte hetzen vor dem Chefkoch
 Eckige Tische schätzen stets Kakteen
 Kesse Sätze hassen heftig das Herz
 Kaputte Köpfe schaffen plötzlich Pfiff
 Kecke Füchse ziehen auf die Piste
 Fesche Tassen töten heute Scherze
 Hübsche Stoffe schauen durch die Post
 Fitte Kioske pachten hierzu Tabak

processing, auditory feedback of self-generated speech). Finally, we tested whether auditory and somatosensory feedback processing during articulation was sensitive to consonant voicing.

METHODS

Participants

Thirty-two right-handed healthy native German speakers (16 females; range 21–30 years, mean age: 25.5) participated in this 3T fMRI study. All had normal or corrected-to-normal vision, no hearing impairments, no known history of neurological disease, or contraindications for fMRI. Participants were paid and written informed consent was obtained according to the rules of the local ethics committee who also approved this study.

Stimuli and Design

This study used 80 declarative semantically nonsense German sentences (10 syllables per sentence) of a fixed syntactic structure (adjective–subject–verb–adverb/preposition–object, Table I). Each sentence was presented once in any of the four experimental conditions. We investigated both listening to and reading sentences. Reading involved either covert or overt reading or articulating sentences without phonating, a condition called mouthing, which lacks auditory output and thus auditory feedback. We expected condition contrasts to reveal correlates of self-produced speech compared with speech perception of pre-recorded sentences of a male

speaker as well as to identify activations associated with inner speech in the mouthing or covert reading condition.

The sentences varied in the amount of phonologically voiceless and voiced consonants while overall consonant and vowel frequency was kept constant. Voiced consonants were sonorants or phonologically voiced obstruents, voiceless consonants were oral stops and fricatives. The percentage of voiceless consonants out of all consonants ranged from 0% to 87%, mean 40% across sentences.

The words of the sentences were mono-, bi-, and trisyllabic. Following the rules of German language, the phonetic modulation regarding voicing went along with increasing numbers of CV syllables. Word frequency and number of alliterations did not systematically co-vary with voicing ($P > 0.05$).

The study design consisted of 4 blocks of 13 minutes and 40 seconds each, resulting in a total acquisition time of 54 minutes and 40 seconds. Each block contained all conditions, and short breaks were introduced in between for the participants' benefit. The trials began with the presentation of a visual cue indicating which task should be performed once a sentence was presented 2–4 seconds later. The association between cue and task was trained prior to the experiment. A triangular cue indicated that the subsequent sentence had to be read out loud (overt reading); a square indicated to read it silently (covert reading); the circle designated that the sentence should be mouthed, while the rhomboidal cue informed participants that they will listen to a pre-recorded sentence. To control for visual input, the auditory stimulation in the listening condition was accompanied by visual presentation of word-like arranged Wingding characters. The sentences, as the sentence-like arranged winding-characters, were presented on a single screen for 3 seconds. Each trial was followed by a jittered baseline during which participants fixated a point for 2–6 seconds (mean 4 seconds). Conditions and sentence identity (more or less voiced or voiceless consonants) were randomized.

The auditory sentences were recorded in a professional recording studio by a male native-German-speaker and presented via MR-compatible headphones (MRConfon, Magdeburg, Germany). Visual stimuli (randomized sentences or sentence-like arranged wingding characters) were projected using Presentation software (Neurobehavioral Systems, Albany, CA) in white on black background on a screen viewed through a head-coil mounted mirror. Participants were instructed to read each sentence as soon as it appeared on the screen or to watch the Wingding characters from left to right while listening to the pre-recorded sentences.

Acoustic Analyses

The pre-recorded auditory sentences were analyzed acoustically for parametric correlations with the imaging data (see below). For generalization purposes, we not only analyzed the male speaker's pre-recorded sentences, but also the recordings of 16 additional women who were

recorded acoustically in a sound proofed room while they read the 80 sentences in randomized order. Such a procedure was chosen, because the recordings of the participants' utterances in the scanner did not allow for a proper acoustic analysis even after scanner noise of the constant acquisition has been filtered out (see below). Split half reliability was high, indicating inter-individual variability was not critical for the studied parameters. The women were instructed to pause between the sentences in order to read and prepare the upcoming sentence and to avoid declination effects from one sentence on another. They read in their own speech rate and speaking style without any particular instructions. The onset and offset of each sentence was labeled manually. For each sentence, data were resampled to 11 kHz and a number of acoustic measures were obtained over the course of the whole sentence, using a time step of 5ms, a window length of 25 ms using a Praatconsole (version 5.3.53, www.praat.org). First, we obtained the fundamental frequency (hereafter F0). The frequency range was set to 80–250 Hz for the male and to 120–400 Hz for the female speakers to avoid octave jumps. All values were additionally visually inspected. The occurrence of phonation was defined as the inverted number of missing F0 values as a percentage of all F0 values. We additionally extracted Sentence duration in ms, Intensity values in dB, and the spectral Center of Gravity (COG) of the entire frequency range in Hz. The COG is an acoustic measure of how high the frequencies in a spectrum are on average. It can be obtained for all sounds (in contrast, e.g., to measuring formants for vowels only). For voiceless sibilants this measure is relatively high while it is much lower in vowels or sonorants [Tabain, 2001, please see Figure 1 for an example of occurrence of F0 and spectral COG in two exemplary sentences]. All F0, Intensity and COG values were averaged over the entire sentence period.

For each sentence, the ratio between the number of voiceless to that of voiced consonants was calculated. Increasing the amount of voiced consonants in the sentences expectedly increased the occurrence of phonation, as revealed by a Spearman ranked correlation using R, version 3.2.3 (females: $r = 0.75$, $P < 0.001$, male: $r = 0.85$, $P < 0.001$), and the mean intensity (females: $r = 0.75$, $P < 0.001$, male: $r = 0.79$, $P < 0.001$), and reduced the mean COG (females: $r = -0.76$, $P < 0.001$, male: $r = -0.81$, $P < 0.001$). Sentence duration did not co-vary with voicing (both $P > 0.15$). Since the occurrence of phonation goes hand in hand with changes in intensity and spectral properties, various acoustic measures may reveal partially redundant information. A Spearman ranked correlation analysis revealed an expectedly high correlation between the parameters: COGmean and INTmean: $r = -0.62$, $P < 0.001$, COGmean and occurrence of phonation, $r = -0.83$, $P < 0.001$, and occurrence of phonation and INTmean, $r = 0.665$, $P < 0.001$. We, therefore, used occurrence of phonation as a single parameter for correlation with fMRI data (see below).

fMRI Data Acquisition

A 3T whole body scanner (Magnetom Trio; Siemens, Erlangen, Germany) equipped with a 1-channel head coil was used to constantly acquire gradient-echo T2*-weighted transverse echo planar images with BOLD contrast. In total, 1632 volumes of an echo-planar imaging sequence (EPI, voxel size = $3.0 \times 3.0 \times 3.0 \text{ mm}^3$, 33 slices to cover the whole brain, TE = 30 ms, TR = 2000 ms) were acquired. The sequence parameters of this continuous acquisition were previously shown to be safe and efficient to study brain activity during 3 seconds of overt speech production [Preibisch et al., 2003]. Participants were instructed to move as little as possible; foam cushions additionally stabilized the head. The participants' utterances were recorded using an MR-compatible microphone (MRConfon, Magdeburg, Germany) placed close to their mouths. To confirm the absence of neurological abnormalities, a high-resolution T1 weighted structural scan was obtained (magnetization prepared rapid acquisition gradient echo, 144 slices, 1 slab, TR 2250 ms, TE 2.6 ms, voxel size $1.0 \times 1.0 \times 1.0 \text{ mm}^3$).

Behavioral Measures during Scanning

The periodic scanner noise of the continuous EPI acquisition was filtered out from audio recordings using an audio processing software (Audacity 2.0.3) and recordings were analyzed to control for task performance. Residual auditory signals were checked and contrasted against the time course of the Presentation logfile. Overall 9 mouthing trials showed minimal articulation-related auditory effects but phonation was not observed. These trials were thus not excluded from analyses. All overt reading trials were correctly performed and no covert trials showed signs of phonation.

fMRI Data Analysis

The voxel-based time-series was analyzed using SPM8 (Statistical Parametric Mapping, www.fil.ion.ucl.ac.uk/spm/). The standard parameters of SPM were used to preprocess the images (realignment, normalization of EPI images to the standard Montreal Neurological Institute EPI template and smoothing with an 8 mm full-width at half-maximum isotropic Gaussian kernel).

Event-related signal changes were modeled separately for each participant using a General Linear Model as implemented in SPM8. The regressors of interest in the individual SPM matrix included four regressors that modeled task-related processing during the 3 seconds execution phase (reading overtly, mouthing while reading, reading covertly, and listening while watching Wingding characters) that were accompanied by four cue-related regressors and six regressors containing the scan-to-scan realignment parameters modeling movement-related artifacts. Movement did not exceed 1.5 mm in translation or rotation. Voicing was studied by including a parametric correlation for each of the four condition-specific regressors during the execution

phase. The occurrence of phonation of each individual sentence (see above) was used as parametric modulator. Regressor-specific activations were identified by convolving the time course of the expected neural signal with a canonical hemodynamic response function. Data were corrected for serial autocorrelations and globally normalized. The resulting model was fitted to the data which yielded voxel-wise beta values containing the individual parameters estimates. Statistical analysis was performed on the group level.

For the second-level group analyses, the individual regressor-specific beta images of the four blocks were entered into a repeated-measures ANOVA. The design matrix contained the condition factor (8 levels: overt reading, mouthing while reading, covert reading, listening all independent of occurrence of phonation, and the parametric modulation by occurrence of phonation separately for each condition) and the participant factor (32 levels). We applied a non-sphericity correction for variance differences across regressors. To detect voxels that correlated either positively or negatively with the occurrence of phonation independently of condition and thus reflect a parametric main effect of occurrence of phonation we investigated the conjunction across conditions by testing the global null hypothesis for positively and negatively weighted parametric modulators separately [Friston et al., 2005]. Such an analysis identifies consistent and joint parametric modulation across conditions. To ascertain significant parametric modulation in each individual condition, we additionally investigated 5 mm spheres around the conjunction result for such effects [small volume correction (SVC), Friston et al., 2005].

Furthermore, we report condition-related effects independent of voicing (contrasted against silent baseline in Fig. 3 and against each other in Fig. 4 and Table II). We specifically explored the contrasts overt reading > mouthing to identify correlates of phonation and consecutive auditory feedback processing and the contrast mouthing > covert reading in order to detect sensorimotor processing-related activations. Auditory cortex suppression during overt reading was investigated by contrasting sentence listening against overt reading. The contrast overt versus covert reading has been intensively investigated in previous studies [Kell et al., 2011; Keller and Kell, 2016] and is not reported here.

All results are thresholded at $P < 0.05$, family wise error (FWE) corrected for multiple comparisons on the voxel level. Coordinates of activations are given in the Montreal Neurological Institute (MNI) space. Brodmann areas corresponding to the activations were identified using probability maps from the Anatomy Toolbox for SPM [Eickhoff et al., 2005] and the Talairach atlas [Lancaster et al., 2000].

To investigate voicing effects within auditory and somatosensory feedback-related activations, we additionally tested for a parametric modulation in the overt reading and mouthing while reading condition separately ($P < 0.05$, FWE corrected in the volume of the condition specific activation).

Regional lateralization of condition and voicing-specific effects were tested by statistically comparing the individual

contrast images with their flipped versions [flipped around the midline, see Josse et al., 2009; Kell et al., 2011; Keller and Kell, 2016; Pichon and Kell, 2013] using two-sample *t*-tests. The resulting lateralization images were thresholded at $P < 0.05$, FWE corrected for multiple comparisons within the volume of the condition or modulation-specific activation at $P < 0.05$, FWE corrected. This procedure tests for regional lateralization by comparing activity voxel-wise with homologue voxels statistically [Seghier, 2008]. Of note, the flipping procedure could potentially evoke spurious lateralization due to structural asymmetries, particularly in the posterior STG [Watkins et al., 2001]. Consequently, homotopes could potentially map imperfectly onto each other. This should result in a mirror lateralization image for the two homotopes. We, thus, carefully checked all lateralization analyses for close-by left- and right-lateralizations but found none.

To investigate whether the observed voicing effect also affected functional lateralization on the lobar or hemispheric level globally, we additionally subjected the relevant contrasts to a threshold-independent calculation of voxel value-lateralization indices in the temporal lobe or hemisphere by using the LI toolbox [Wilke and Lidzba, 2007]. Voicing did not significantly affect global functional lateralization.

Voicing could also modulate the functional integration of voice-sensitive brain regions in ipsi- and contralateral cerebral networks independent of lateralization. In this case voicing would not only affect local processing but also change the way the information is integrated in the speech network. Therefore, we additionally investigated whether voicing modulated the functional connectivity of the observed voicing-sensitive brain regions. In exact analogy to Keller and Kell [2016], we calculated psychophysiological interactions

using the time courses extracted from voicing-sensitive brain regions using the parametric modulator “occurrence of phonation” as psychological variable. For methodological details, please see Keller and Kell, 2016. Voicing affected only regional activity but did not significantly modulate functional connectivity of voicing-sensitive brain regions ($P > 0.05$).

RESULTS

Acoustic Analyses

Sentences rich in voiceless consonants showed much larger spectral changes as well as much larger temporal modulations of the amplitude envelope (see Fig. 1 for representative examples). The lack of fundamental frequency

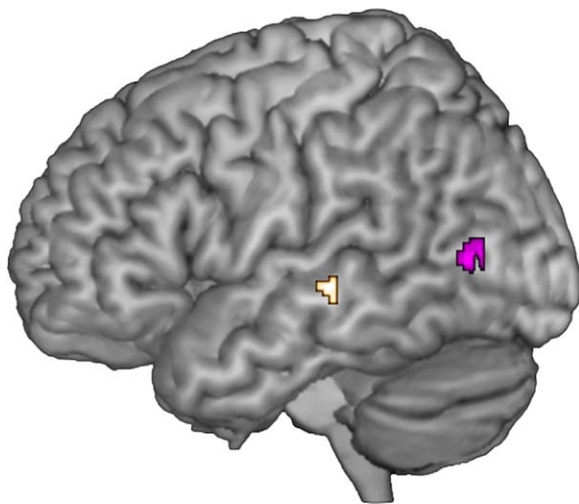


Figure 2.

Condition-independent parametric modulation by occurrence of phonation. Activity increase with voicing (gold) and activity increase with increasing numbers of voiceless consonants (purple) at $P < 0.05$, FWE corrected.

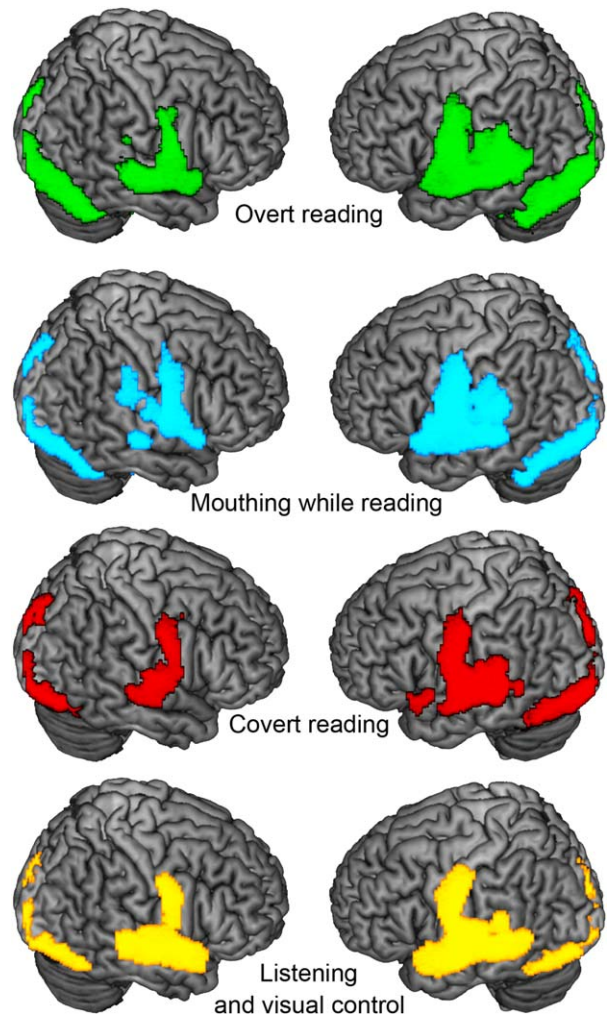


Figure 3.

Condition effects. All activations (conditions contrasted against silent baseline) are depicted at $P < 0.05$, FWE corrected. Green: Overt reading, Blue: Mouthing while reading, Red: Covert reading, Yellow: Listening while viewing wingding characters from left to right.

during voiceless consonants produced a marked increase in the spectral center of gravity (COG). In parallel, the lack of phonation during voiceless consonants induced numerous acoustic onsets in sentences with more voiceless consonants compared with sentences with more voiced consonants, during which phonation occurred almost constantly. Thus, the experimental manipulation had both spectral (lack of F0) and temporal consequences (faster temporal modulation of the speech envelope).

fMRI Analyses

Independent of condition, the more voiced consonants a given sentence contained, the stronger the activity was in the lower bank of the left superior temporal sulcus (STS, MNI $-64 -24 -2$, $Z = 5.28$, $P = 0.01$, FWE corrected, golden cluster in Fig. 2). All conditions showed this effect when tested separately (all $P < 0.05$, SVC), except for the listening condition in which parametric modulation was subthreshold ($P = 0.097$) due to ceiling beta values. The

right STS also showed this condition-independent parametric modulation by occurrence of phonation, but slightly subthreshold (MNI $68 -20 -2$, $Z = 4.88$, $P = 0.064$). Consequently, this parametric effect was not lateralized ($P > 0.05$). Voicing did not modulate the functional connectivity of these brain regions ($P > 0.05$).

The less a sentence was phonated (consisted of more voiceless consonants) the stronger activity was in the left posterior middle temporal gyrus (MTG) independent of condition (MNI $-54 -70 6$, $Z = 6.29$, $P < 0.001$, FWE corrected, purple cluster in Fig. 2). This effect was observed in every single condition at $P < 0.05$ (SVC) and was left-lateralized ($Z = 2.72$, $P = 0.02$, SVC). This lateralization occurred only on the regional level given that it was not observed when functional lateralization was studied on the level of the entire temporal lobe or hemisphere. Voicelessness did not significantly change the functional integration of this region in the speech network (PPI analyses). Even when studying the overt reading or the mouthing condition separately, there was no significant voicing-related modulation in the motor cortex (all $P > 0.05$).

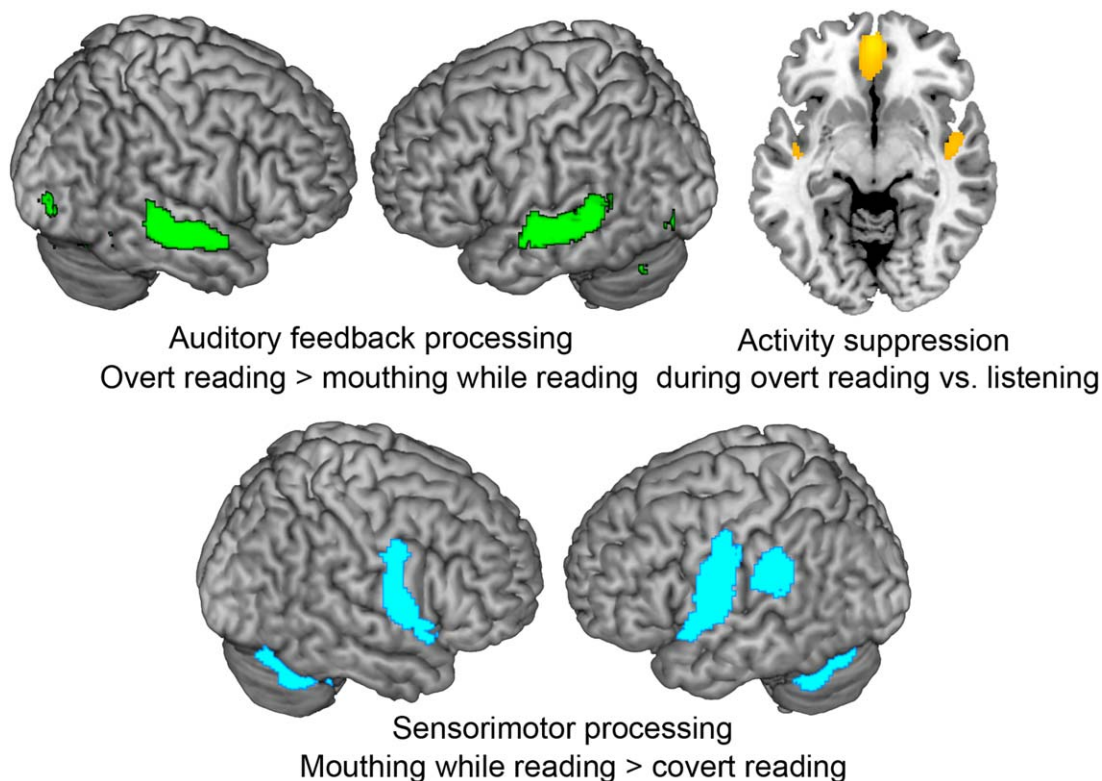


Figure 4.

Condition differences related with sensory feedback processing. Upper panel: The contrast overt reading > mouthing while reading reveals bilateral activations associated with auditory feedback processing (green). Activity in the bilateral anterior auditory cortices and in the ventromesial prefrontal cortex is suppressed

during overt reading compared with listening (yellow). Lower panel: Activations related with sensorimotor processing as revealed by the contrast mouthing while reading > covert reading (blue). Note the left-lateralization in secondary somatosensory cortices. All activations at $P < 0.05$, FWE corrected.

Independent of voicing, the experimental conditions revealed a large bilateral perisylvian, occipitotemporal, and cerebellar-subcortical task-relevant network (see Fig. 3 for the conditions separately). Covert reading and sentence listening activated the bilateral premotor cortex in the context of this experiment. When overt reading was compared with the mouthing condition, there was additional activation along the bilateral posterior and mid portion of the STS and posterolateral cerebellar hemispheres without any lateralization (all $P > 0.05$), reflecting bilateral processing of the auditory feedback signal (Fig. 4, upper left panel and Table II).

This auditory feedback-related activation encompassed the region that activated more strongly for more voiced sentences independent of condition (see above). When studying voicing effects in the overt reading condition separately (the only condition evoking auditory feedback), there was a modulation of STS activity, in the sense that there was more activity for more voiced sentences that became significant only in the left temporal cortex (MNI $-52 -26 -4$, $Z = 4.68$, $P = 0.008$, SVC).

We also observed the well-described relative suppression of auditory cortex activity during self-produced speech when activity is compared with passive listening [Agnew et al., 2013]. Activity in the bilateral planum polare, anterior superior temporal gyrus, and the ventromedial prefrontal cortex (part of the default mode network) was weaker when speech was self-generated during overt reading compared with sentence listening (Fig. 4, upper right panel and Table II). Auditory cortex suppression during overt reading was bilateral (lateralization analysis, all $P > 0.05$). Activity in the anterior temporal cortices and in the dorsal speech processing stream during overt reading was not significantly different from covert reading or mouthing, suggesting these conditions evoked inner speech. Furthermore, mouthing did not show stronger activity when compared with overt reading.

In comparison to covert reading without mouthing, mouthing while reading covertly activated the bilateral inferior Rolandic cortex (sensorimotor cortex associated with articulation), the sensorimotor thalamus, the bilateral cerebellar hemispheres, and the left secondary somatosensory cortex in the anterior portion of the supramarginal gyrus (Fig. 4, lower panel and Table II) more strongly. The left somatosensory activation was the only lateralized activation in this contrast ($Z = 4.42$, $P = 0.011$, SVC). This activation likely reflects somatosensory feedback processing during articulation. There was no significant correlation with occurrence of phonation in this region ($P > 0.05$). Covert reading did not activate any region more strongly than mouthing while reading ($P > 0.05$).

DISCUSSION

In all conditions independent of input modality, a positive correlation with occurrence of phonation was found in

voice-sensitive regions in the bilateral STS. This represents an unsurprising consequence of increased voicing in the overt reading and listening condition, but the same relationship was observed during covert reading with or without mouthing. This confirms inner voice-related processing in the bilateral STS during covert reading [Perrone-Bertolotti et al., 2012; Yao et al., 2011]. This observation stood in contrast to the left-lateralized activity in left posterior middle temporal gyrus (MTG), which increased with increasing numbers of voiceless consonants in the processed sentences, independent of condition. During overt reading, auditory feedback-related processing was sensitive to consonant features and concomitant acoustic changes while voicing did not significantly modulate somatosensory feedback processing during mouthing.

Lateralization

There are two possible ways lateralized activations are classically interpreted. One interpretation conceives lateralization as a result of fundamental processing differences between the hemispheres. Such fundamental processing differences should result in a specialized region in one hemisphere whose lesion induces a drastic loss of function. One may thus expect a lesion of the left posterior MTG to result in a specific processing deficit for voiceless consonants. This is clearly not observed in neurological patients. Rather, lateralized regional activity related with an experimental factor embedded in an overall bilateral network may index subtle preferential processing that could be useful when hemispheres divide processing load between them. When an activation is lateralized, this does not imply that the contralateral homotope does not contribute to processing. It simply suggests that a region contributes significantly more to processing than its homotope. Such subtle processing preferences could result in actual functional specialization, should learning reinforce such hemispheric asymmetries by strengthening ipsilateral network connectivity and disengaging the contralateral homotope from processing [Keller and Kell, 2016; Minagawa-Kawai et al., 2011].

Our study focuses on sentence reading, a condition that activates both hemispheres. It is the condition-dependent regional lateralization within this overall bilateral network that provides valuable information on hemispheric processing preferences.

Inner Speech during Reading Codes Phonetic Detail

Independent of condition, the degree of voicing modulated activity in the temporal lobe, but not in the motor cortices. Such a finding was expected since laryngeal mechanisms are involved in the production of both voiced and voiceless consonants. The latter was not even observed when articulatory routines were executed during overt reading or

mouthings. This suggests that voicing effects can be detected more easily when focusing on auditory features, even during production. Acoustically, increased voicing coincided with longer episodes of fundamental frequency and larger mean intensity as well as lower average spectral COG values. Decreased voicing instead went along with higher spectral COG and larger temporal modulation of intensity. Yet, neural voicing effects were observed in auditory association cortices, even in the absence of any auditory input in the inner speech conditions. Our results thus provide empirical evidence that inner speech during reading yields similar phonetic detail compared with overt speech, at least with regard to the voicing contrast. These findings also demonstrate that inner speech relies both on covert motor processing and on an emulation of the sensory consequences of articulation [Hickok, 2012].

Bilateral Voice Processing

Sentence reading elicited an overall bilateral pattern of activation. Increased voicing, even when emulated during inner speech, activated voice-selective temporal regions without significant lateralization. The bilateral central STS is highly voice selective during auditory processing [Belin et al., 2000] and represents auditory feedback-related intensity changes (the amplitude envelope) during overt sentence reading [Pichon and Kell, 2013]. These results demonstrate that inner speech during reading simulates the amplitude envelope. The degree of voicing did not change functional integration of these voice-sensitive regions in speech networks. Such finding points toward a local analysis of voicing that constantly informs speech processing.

A bilateral activation related to voice processing is predicted by speech processing frameworks focusing either on the spectral [Ivry and Robertson, 1999] or temporal domain [Poeppel, 2003]. Ivry and Robertson propose that lower spectral frequencies (e.g., in the range of the fundamental frequency) can be decoded by the right auditory cortex while Poeppel's Asymmetric Sampling in Time hypothesis posits that the right auditory cortex may contribute to speech perception by analyzing longer temporal windows, a prerequisite for analyzing sentence-level prosody/amplitude envelope. Left-lateralization in the two theories pertains to processing of either relative higher spectral frequencies [Ivry and Robertson, 1999] or shorter temporal integration windows [Poeppel, 2003]. Whatever the exact lateralizing factor, our results suggest that the proposed lateralization rule generalizes to simulated speech during the inner speech conditions. This does not exclude a sensory origin of speech processing lateralization [Kell et al., 2011], but rather suggests that inner speech involves simulations of sensory consequences of speaking that follow the same lateralization pattern.

Preferential Phonetic Processing in the Left Posterior Temporal Lobe

The more voiceless consonants a sentence contained, the more left-lateralized processing occurred in the posterior MTG, independent of condition and of auditory input. Voicelessness only affected activity regionally in the posterior MTG and not globally in the left temporal lobe or hemisphere. Moreover, voicelessness did not change functional integration of this region in the speech network. This confirms that voicing constitutes a speech feature that is constantly monitored and integrated. In agreement with the prediction of the Double Filtering by Frequency hypothesis [Ivry and Robertson, 1999], voiceless consonants activate the left auditory cortex more strongly than voiced consonants during auditory syllable discrimination [Jancke et al., 2002]. Yet, the observed association between processing of voiceless consonants and brain activity in this study reached well beyond auditory cortices involved in spectral processing. The left posterior MTG lies at the ventro-caudal border of regions involved in phonological processing of auditory speech [Hickok and Poeppel, 2007; Vigneau et al., 2006] and has been associated with phonological processing during reading [Price, 2012]. Given that reading is acquired years after speech comprehension in human speech acquisition, this effect in phonology-related cortices could potentially represent a downstream consequence of left-lateralized preferential processing of voiceless consonants (i.e., higher relative spectral frequencies) in auditory cortices [Jancke et al., 2002]. In comparison to the study by Jancke and colleagues, we acquired EPI volumes continuously to increase the sampling rate at the expense of continuous scanner noise. This is known to mask auditory responses [Shah et al., 2000; Talavage et al., 1999] and could explain our negative finding in primary auditory cortices. Yet, the posterior MTG activation, as well, carried phonetic information in the sense that it was more strongly activated when the number of voiceless consonants increased.

The experimental manipulation of consonant features resulted not only in spectral but also in temporal changes of sentence properties. Indeed, the higher the spectral COG, the stronger the temporal intensity modulation of the sentences. Thus, the left-lateralization observed in the posterior MTG could also be interpreted as a consequence of lateralized temporal processing underlying both overt and inner speech. The topographical dissociation between brain areas sensitive to increased mean intensity in the bilateral STS and left-lateralized MTG sensitive to sentences that vary greatly in intensity (sentences rich in voiceless consonants) is suggestive of temporal factors contributing to lateralization in MTG. Although this activation was also observed in the inner speech conditions, we consult speech perception studies to interpret this effect.

Neural oscillations at high temporal modulation frequencies (in the gamma range) reliably and quickly track

TABLE II. Condition contrasts

Contrast	Anatomical region	MNI coordinates (x y z)	Z (P) value (FWE corr.)
Overt reading > Mouthing while reading	L superior temporal sulcus	-60 -24 -2	Inf (0.000)
	L superior temporal sulcus	-60 -36 2	7.45 (0.000)
	R superior temporal sulcus	50 -20 -8	Inf (0.000)
	R superior temporal sulcus	60 -28 0	Inf (0.000)
	L temporo-occipital junction	-42 -54 -10	7.72 (0.000)
	R temporo-occipital junction	42 -54 -10	5.92 (0.001)
	Cerebellar vermis	0 -66 -34	Inf (0.000)
	L posterior cerebellar hemisphere	-18 -72 -38	7.46 (0.000)
	R posterior cerebellar hemisphere	16 -76 -40	7.22 (0.000)
Mouthing > Covert reading	L articulatory motor cortex	-56 -6 14	Inf (0.000)
	R articulatory motor cortex	62 0 12	7.45 (0.000)
	L primary motor cortex	-54 -8 34	6.62 (0.000)
	R primary motor cortex	58 -2 42	6.56 (0.000)
	L ventral premotor cortex	-60 4 4	6.40 (0.000)
	R ventral premotor cortex	52 14 -6	6.45 (0.000)
	L anterior supramarginal gyrus	-60 -40 30	6.17 (0.000)
	L anterior supramarginal gyrus	-62 -32 36	5.51 (0.000)
	L ventrolateral thalamus	-12 -16 8	6.85 (0.000)
	R ventrolateral thalamus	18 -14 12	5.30 (0.002)
	L superior cerebellar hemisphere	-16 -64 -22	Inf (0.000)
	R superior cerebellar hemisphere	18 -62 -24	Inf (0.000)
	L inferior medial cerebellar hemisphere	-12 -66 -50	7.75 (0.000)
	R inferior medial cerebellar hemisphere	12 -68 -50	Inf (0.000)
	L medial planum polare	-44 -12 -8	5.07 (0.000)
	R medial planum polare	46 -6 -10	6.3 (0.001)
Listening > Overt reading	L anterior superior temporal gyrus	-40 14 -26	Inf (0.000)
	R anterior superior temporal gyrus	44 16 -20	Inf (0.000)
	R anterior temporal pole	38 18 -30	Inf (0.000)
	Ventromedial prefrontal cortex (DMN)	0 52 -6	Inf (0.000)
	Ventromedial prefrontal cortex (DMN)	0 46 26	Inf (0.000)

the fast modulation of the speech envelope [Nourski et al., 2009; Zion Golumbic et al., 2013] and thus, provide ideal neural substrates to facilitate analyses of fast changes in the relative high spectral frequencies underlying consonant identification. During sentence processing, auditory cortices entrain to both low and high temporal modulation frequencies. Temporal speech-brain alignment occurs both in the delta/theta and gamma range and is thought to facilitate perception [Giraud and Poeppel, 2012]. Of note, our phonological manipulation resulted in increased intensity modulation in the theta range as sentences contained more voiceless consonants (see Fig. 1 for a representative example). During processing of sentences rich in voiceless consonants, consonant identification could benefit from enhanced cross-frequency interactions of temporal modulation frequencies due to increased power in low temporal modulation frequencies. Such a cross-frequency coupling has previously been shown to be left-lateralized and to also occur in the posterior MTG during auditory sentence processing [Gross et al., 2013].

Whether the underlying causes for lateralization are spectral or temporal, the effect was not only observed during sentence listening and overt reading, but also during covert reading and mouthing while reading. These

two latter conditions involve transformations of orthography into phonology without auditory input and evoke the impression of inner speech [Perrone-Bertolotti et al., 2014]. Here, we show that this transformation occurs at least at a level of consonant features. We cannot claim inner speech in general provides such detail, but inner speech during reading does so, even for sentence reading that is assumed to rely much less on grapheme/phoneme relations than single word reading, given the additional contextual information. The detailedness of inner speech during reading resolving even consonant features could facilitate reading acquisition and represents a prerequisite for a facilitatory role of inner speech in grapheme/phoneme mapping.

Left-Lateralized Speech-Related Somatosensory Feedback Processing

Overall, the conditions of interest activated bilateral occipital and perisylvian regions together with basal ganglia, thalamus, and cerebellum. The bilateral premotor cortex activated not only in the overt speech, but also in the covert reading and passive listening conditions. For both

latter conditions, motor cortex involvement has been shown, albeit with varying degrees of involvement between studies. We noted a clear influence of experimental conditions tested in parallel to a reference condition throughout our studies [see Kell et al., 2011; Keller and Kell, 2016; Pichon and Kell, 2013]. It is likely that the mouthing condition and the associated training prior to scanning induced a focus on sensorimotor speech aspects even in the listening and covert reading condition. Furthermore, the low semantic content of the sentences may have evoked stronger phonological processing.

Nevertheless, clear differences between conditions were observed. Mouthing involves articulatory movements independent of phonation and evokes somatosensory feedback only. The increased mouthing-related activity in the bilateral cerebellum and bilateral ventral Rolandic cortex together with the left-lateralized activity in the secondary somatosensory cortex extending into the temporo-parietal junction can likely be attributed to the increased sensorimotor processing demands associated with mouthing while reading. Of note, the left-lateralization as defined here only implies stronger activation in the left compared with the right hemisphere and does not exclude activation of the right hemispheric homologues. Bilateral somatosensory feedback processing in inferior parietal/parieto-opercular regions was previously shown for movements of the articulators that do not have auditory targets [Dhanjal et al., 2008] while our results imply that the left secondary somatosensory cortex is preferentially active when sentences are mouthed [see also Agnew et al., 2013]. This suggests a context-dependent lateralization of somatosensory feedback processing. The speech-induced lateralization of somatosensory feedback processing could either point to preferences for processing speech-relevant aspects of the sensory input in the left somatosensory cortex [Ivry and Robertson, 1999], to left-lateralized auditory-somatosensory interactions, or to an interaction with the left-lateralized language system. Interestingly, the mouthing-related lateralization of somatosensory feedback processing was not paralleled by the lateralization of articulatory processes in the motor cortices, arguing for a special role of sensory processing in the lateralization of articulation [see also Kell et al., 2011]. The DIVA model proposes the right-lateralized premotor cortex adjusts motor output to perturbations of somatosensory feedback that evoke bilateral activations of somatosensory cortices [Golfingopoulos et al., 2011]. Such a right-lateralization of motor adaptations could result from the use of sudden unexpected feedback perturbations. Such acute online adaptations may very well differ from physiological monitoring processes during unperturbed continuous speech in which error signals would require slower and subtler adaptations to performance drifts.

Voicing did not significantly modulate the inferior parietal activity associated with articulation-related somatosensory feedback. Together with the lack of voicing effects in the motor cortices, this may point to preferential

processing of place of articulation in the articulatory sensorimotor system [Bouchard et al., 2013].

Bilateral Speech-Related Auditory Feedback Processing

The brain uses auditory feedback for different aspects of speech motor control. Such auditory feedback is monitored to adjust speech intensity [Lombard, 1911], maintain fluency [Yates, 1963], and tune the speech motor system, as exemplified by an articulatory decline in the hearing impaired [Waldstein, 1990]. While most adaptations to changes in auditory feedback occur automatically, the DIVA model suggests the cortex contributes at least to the latter. In line with the predictions of the DIVA and hierarchical state feedback control models, self-generated or pre-recorded auditory input during overt reading and sentence listening further increased bilateral superior temporal activity compared with the covert reading conditions. In contrast, the additional auditory input did not alter the activity in the dorsal speech processing stream compared with the covert reading conditions. This suggests external auditory feedback processing in bilateral anterior temporal cortices on top of computations along both ventral and dorsal speech processing streams during the covert conditions evoking inner speech [see also Keller and Kell, 2016]. This does not exclude auditory feedback processing in the dorsal stream, but suggests a dissociation between the sensitivity toward external auditory stimulation in ventral and dorsal pathways. Since this study and Agnew et al. [2013] show auditory feedback-related activations in mid and anterior portions of the bilateral superior temporal cortices and both used sentences as auditory stimuli, a reasonable interpretation could be the existence of a monitoring mechanism on the sentence level. The sentence-level auditory processing has been associated with bilateral anterior parts of the superior temporal cortices [Humphries et al., 2005]. In contrast, manipulations of formant frequencies in self-generated monosyllabic words affected the dorsal speech processing stream more strongly [Tourville et al., 2008].

The results of Agnew et al., this report, and one of our former studies [Pichon and Kell, 2013] suggest that the superior portions of the mid and anterior temporal lobe also monitor suprasegmental aspects of self-generated speech including prosodic information. Indeed, auditory feedback-related activity in the superior temporal cortex was sensitive to the temporal dynamics of the speech envelope. Importantly, this was also observed in the inner speech conditions. Our results thus support the notion that inner speech is based upon internal feedback loops [Guenther and Hickok, 2015].

Limitations

This study cannot dissociate whether the observed voicing effect is a consequence of auditory processing alone or

whether it results from an interaction with the language system [Scott and McGettigan, 2013]. Even though left-lateralized processing of voiceless consonants was also observed during inner speech, a close relationship of activity in these conditions with acoustic features was found. It could represent a consequence of a left auditory preference for spectro-temporal speech analyses, resulting in a left-hemisphere advantage for language processing in general [Minagawa-Kawai et al., 2011]. Studies carried out on congenitally deaf people could help answering the question on whether these processing preferences are modality specific or domain general.

We detected voicing effects in auditory association cortices involved in phonological processing. This acoustic feature may go hand in hand with syllable complexity on the phonological level. We cannot disentangle whether the differences arise on one level or the other of the theoretical constructs. Given their close relationship, we question the relevance of such dissociation for consonant processing.

Finally, this study was not suited for investigating effects of place of articulation. Such processing involves the left-lateralized dorsal speech processing stream [Murakami et al., 2015].

CONCLUSIONS

Our results provide evidence that inner speech entails both covert motor processing and an emulation of the sensory consequences of articulation. The direct relationship between modulation of temporal activity during sentence processing and acoustic parameters even in the conditions lacking auditory input suggests inner speech during reading represents detail as fine as consonants' voicing features. Such detail is required in order to grant inner speech a facilitatory role in reading acquisition by supporting grapheme/phoneme transformations.

Our results confirm models of inner speech that incorporate internal fronto-temporal loops. Yet, both during inner and overt speech, the degree of voicing affected posterior and anterior regions of the temporal lobe differently. This suggests the DIVA or hierarchical state feedback control models could be improved if auditory targets are sub specified according to different acoustic dimensions. Our results imply that bilateral computations of the acoustic envelope in more anterior temporal cortices aid analyses of covert or overt prosodic features. In contrast, higher temporal modulation rates and/or higher spectral frequencies of the very same emulated or overt speech signal activated more posterior left-lateralized speech representations. Together with observations of place of articulation-related processing in the left-lateralized dorsal speech processing stream [Murakami et al., 2015], these results are suggestive of an anterior-posterior gradient in temporal processing in the temporal lobe. More anterior auditory cortices associated with ventral stream processing could process slower modulations of the speech signal like prosody while posterior auditory cortices

giving rise to the dorsal speech processing stream may primarily be sensitive to the fast acoustic changes [Baumann et al., 2015].

During overt reading, voicing features modulated left superior temporal cortex activity in an overall bilateral superior temporal network related to auditory feedback processing. Left-lateralized somatosensory feedback and bilateral motor processing was insensitive to voicing. This suggests voicing is primarily monitored in the auditory rather than in the somatosensory feedback channel and implies different weighting of the proposed separate fronto-temporal loops in speech models for different aspects of speech motor control.

ACKNOWLEDGMENTS

We would like to thank Anja Pflug for scripting. The authors declare there is no potential conflict of interest related with this publication.

REFERENCES

- Agnew ZK, McGettigan C, Banks B, Scott SK (2013): Articulatory movements modulate auditory responses to speech. *Neuroimage* 73:191–199.
- Alderson-Day B, Fernyhough C (2015): Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychol Bull* 141:931–965.
- Arnold C, Gehrig J, Gispert S, Seifried C, Kell CA (2013): Pathomechanisms and compensatory efforts related to Parkinsonian speech. *NeuroImage Clin* 4:82–97.
- Arsenault JS, Buchsbaum BR (2015): Distributed neural representations of phonological features during speech perception. *J Neurosci* 35:634–642.
- Baumann S, Joly O, Rees A, Petkov CI, Sun L, Thiele A, Griffiths TD (2015): The topography of frequency and time representation in primate auditory cortices. *eLife*, 4:e03256.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Binder JR, Desai RH, Graves WW, Conant LL (2009): Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Boemio A, Fromm S, Braun A, Poeppel D (2005): Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature neuroscience*, 8:389–395.
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013): Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332.
- Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G (2011): Conduction aphasia, sensory-motor integration, and phonological short-term memory—an aggregate analysis of lesion and fMRI data. *Brain Lang* 119:119–128.
- Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B (2014): Sensory-motor transformations for speech occur bilaterally. *Nature* 507:94–98.
- Conrad B, Schönle P (1979): Speech and respiration. *Arch Psychiatr Nervenkrankheiten* 226:251–268.
- Corley M, Brocklehurst PH, Moat HS (2011): Error biases in inner and overt speech: evidence from tongue twisters. *J Exp Psychol* 37:162–175.

- Dhanjal NS, Handunnetthi L, Patel MC, Wise RJ (2008): Perceptual systems controlling speech production. *J Neurosci* 28: 9969–9975.
- De Nil LF, Kroll RM, Kapur S, Houle S (2000): A positron emission tomography study of silent and oral single word reading in stuttering and nonstuttering adults. *JSLHR* 43:1038–1053.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005): A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335.
- Friston KJ, Penny WD, Glaser DE (2005): Conjunction revisited. *Neuroimage* 25:661–667.
- Giraud AL, Poeppel D (2012): Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH (2011): fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage* 55:1324–1338.
- Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013): Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol* 11:e1001752.
- Gruber O, von Cramon DY (2003): The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks. *Neuroimage* 19:797–809.
- Guenther FH, Hickok G (2015): Role of the auditory system in speech production. *Handb Clin Neurol* 129:161–175.
- Hall TA (2000): *Phonologie. Eine Einführung*. Berlin & New York: Walter de Gruyter.
- Hickok G (2012): Computational neuroanatomy of speech production. *Nat Rev Neurosci* 13:135–145.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003): Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci* 15:673–682.
- Hickok G, Okada K, Serences JT (2009): Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101:2725–2732.
- Humphries C, Love T, Swinney D, Hickok G (2005): Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum Brain Mapp* 26:128–138.
- Ivry R, Robertson LC (1999): *The Two Sides of Perception*. Cambridge: MIT Press.
- Jancke L, Wustenberg T, Scheich H, Heinze HJ (2002): Phonetic perception and the temporal cortex. *Neuroimage* 15:733–746.
- Jobard G, Crivello F, Tzourio-Mazoyer N (2003): Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage* 20:693–712.
- Jones PE (2009): From ‘external speech’ to ‘inner speech’ in Vygotsky: A critical appraisal and fresh perspectives. *Lang Commun* 29:166–181.
- Josse G, Kherif F, Flandin G, Seghier ML, Price CJ (2009): Predicting language lateralization from gray matter. *J Neurosci* 29: 13516–13523.
- Kell CA, Morillon B, Kouneiher F, Giraud AL (2011): Lateralization of speech production starts in sensory cortices—a possible sensory origin of cerebral left dominance for speech. *Cereb Cortex* 21:932–937.
- Keller C, Kell CA (2016): Asymmetric intra- and interhemispheric interactions during covert and overt sentence reading. *Neuropsychologia* Epub ahead of print, 4th April 2016, pii: S0028-3932:30111–30117. doi: 10.1016/j.neuropsychologia.2016.04.002.
- Kentner G, Vasisht S (2016): Prosodic focus marking in silent reading: effects of discourse context and rhythm. *Front Psychol* 7:319.
- Koelsch S, Schulze K, Sammler D, Fritz T, Muller K, Gruber O (2009): Functional architecture of verbal and tonal working memory: an fMRI study. *Hum Brain Mapp* 30:859–873.
- Kohler K (1990): German. *J Int Phonet Assoc* 20:48–50.
- Ladefoged P, Maddieson I (1996): *The Sounds of the World’s Languages*. Cambridge, MA: Oxford Press.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000): Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120–131.
- Loevenbruck H, Baci M, Segebarth C, Abry C (2005): The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *J Neuroling* 18:237–258.
- Lombard E (1911): *Le signe de l’élévation de la voix*. Annales des maladies de l’oreille, du larynx, du nez et du pharynx. Paris: Masson.
- Luo H, Poeppel D (2007): Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.
- Marschall BP (1995): *Einführung in die Phonetik*. Berlin/New York: Walter de Gruyter.
- Mazoyer B, Mellet E, Perchey G, Zago L, Crivello F, Jobard G, Delcroix N, Vigneau M, Leroux G, Petit L, Joliot M, Tzourio-Mazoyer N (2016): BIL&GIN: A neuroimaging, cognitive, behavioral, and genetic database for the study of human brain lateralization. *Neuroimage* 124:1225–1231.
- Mesgarani N, Cheung C, Johnson K, Chang EF (2014): Phonetic feature encoding in human superior temporal gyrus. *Science* 343:1006–1010.
- Minagawa-Kawai Y, Cristia A, Dupoux E (2011): Cerebral lateralization and early speech acquisition: a developmental scenario. *Dev Cogn Neurosci* 1:217–232.
- Murakami T, Kell CA, Restle J, Ugawa Y, Ziemann U (2015): Left dorsal speech stream components and their contribution to phonological processing. *J Neurosci* 35:1411–1422.
- Nooteboom SG (1997): The prosody of speech: melody and rhythm. In: Hardcastle, WJ, Laver J, editors. *The Handbook of Phonetic Sciences*. Massachusetts: MIT. pp. 640–673.
- Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, 3rd, Brugge JF (2009): Temporal envelope of time-compressed speech represented in the human auditory cortex. *J Neurosci* 29:15564–15574.
- Nourski KV, Steinschneider M, Rhone AE, Oya H, Kawasaki H, Howard MA, 3rd, McMurray B (2015): Sound identification in human auditory cortex: Differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings. *Brain Lang* 148:37–50.
- Oppenheim GM, Dell GS (2008): Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106:528–537.
- Oppenheim GM, Dell GS (2010): Motor movement matters: the flexible abstractness of inner speech. *Mem Cogn* 38:1147–1160.
- Perrone-Bertolotti M, Kujala J, Vidal JR, Hamame CM, Ossandon T, Bertrand O, Minotti L, Kahane P, Jerbi K, Lachaux JP (2012): How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *J Neurosci* 32:17554–17562.

- Perrone-Bertolotti M, Rapin L, Lachaux JP, Baciú M, Loevenbruck H (2014): What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behav Brain Res* 261:220–239.
- Pichon S, Kell CA (2013): Affective and sensorimotor components of emotional prosody generation. *J Neurosci* 33:1640–1650.
- Pickett JM (1999): *The Acoustics of Speech Communication. Fundamental, Speech Perception Theory and Technology*. Boston: Allyn & Bacon.
- Poeppl D (2003): The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun* 41:245–255.
- Preibisch C, Raab P, Neumann K, Euler HA, von Gudenberg AW, Gall V, Lanfermann H, Zanella F (2003): Event-related fMRI for the suppression of speech-associated artifacts in stuttering. *Neuroimage* 19:1076–1084.
- Price CJ (2012): A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62:816–847.
- Scott SK, McGettigan C (2013): Do temporal processes underlie left hemisphere dominance in speech perception?. *Brain Lang* 127:36–45.
- Seghier ML (2008): Laterality index in functional MRI: methodological issues. *Magn Reson Imag* 26:594–601.
- Shah NJ, Steinhoff S, Mirzazade S, Zafiris O, Grosse-Ruyken ML, Jancke L, Zilles K (2000): The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. *Neuroimage* 12:100–108.
- Simmonds AJ, Leech R, Collins C, Redjep O, Wise RJ (2014): Sensory-motor integration during speech production localizes to both left and right plana temporale. *J Neurosci* 34:12963–12972.
- Tabain M (2001): Variability in fricative production and spectra: implications for the hyper- and hypo- and quantal theories of speech production. *Lang Speech* 44:57–94.
- Talavage TM, Edmister WB, Ledden PJ, Weisskoff RM (1999): Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Hum Brain Mapp* 7:79–88.
- Taylor J, Rastle K, Davis MH (2013): Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychol Bull* 139: 766–791.
- Titze IR, Schmidt SS, Titze MR (1995): Phonation threshold pressure in a physical model of the vocal fold mucosa. *J Acoust Soc Am* 97:3080–3084.
- Tourville JA, Guenther FH (2011): The DIVA model: A neural theory of speech acquisition and production. *Lang Cogn Process* 26:952–981.
- Tourville JA, Reilly KJ, Guenther FH (2008): Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39:1429–1443.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N (2006): Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30:1414–1432.
- Waldstein RS (1990): Effects of postlingual deafness on speech production: implications for the role of auditory feedback. *J Acoust Soc Am* 88:2099–2114.
- Watkins KE, Paus T, Lerch JP, Zijdenbos A, Collins DL, Neelin P, Taylor J, Worsley KJ, Evans AC (2001): Structural asymmetries in the human brain: a voxel-based statistical analysis of 142 MRI scans. *Cereb Cortex* 11:868–877.
- Wilke M, Lidzba K (2007): LI-tool: a new toolbox to assess lateralization in functional MR-data. *J Neurosci Meth* 163: 128–136.
- Yao B, Belin P, Scheepers C (2011): Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *J Cogn Neurosci* 23:3146–3152.
- Yates AJ (1963): Delayed auditory feedback. *Psychol Bull* 60: 213–232.
- Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013): Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77:980–991.